UPORABA UMETNE INTELIGENCE PRI ODKRIVANJU NOVIH UČINKOVIN

THE USE OF ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY

AVTORJA / AUTHORS:
asist. Matic Proj, mag. farm.
asist. Žan Toplak, mag. farm.

Univerza v Ljubljani, Fakulteta za farmacijo, Katedra za farmacevtsko kemijo, Aškerčeva 7, 1000 Ljubljana

NASLOV ZA DOPISOVANJE / CORRESPONDENCE:
E-mail: zan.toplak@ffa.uni-lj.si

POVZETEK

Razmah umetne inteligence na številnih področjih prodira tudi v farmacevtsko industrijo in akademske laboratorije, ki se ukvarjajo z odkrivanjem novih učinkovin. Nove algoritme za iskanje vzorcev v vedno večjih količinah podatkov že uporabljamo za pomoč pri racionalnem načrtovanju učinkovin. V literaturi so opisali številne primere uporabe na področjih odkrivanja tarč in priprave njihove strukture, priprave virtualnih knjižnic spojin, načrtovanja novih spojin zadetkov, njihovo optimizacijo do spojin vodnic, predkliničnih in kliničnih raziskav. Kljub temu je na tem področju še veliko izzivov.

KLJUČNE BESEDE:
podatkovne zbirke, računalniško podprto načrtovanje učinkovin, strojno učenje, umetna inteligenca

ABSTRACT

The rise of artificial intelligence in many fields is also making its way into the pharmaceutical industry and academic labs engaged in drug discovery. New algorithms for pattern recognition in ever-increasing amounts of data are already being used to support rational design. The scientific literature is replete with reports of applications in the areas of target discovery and structure preparation, virtual compound library design, discovery of new hit compounds and optimisation to lead compounds, preclinical and clinical trials. However, there are still a number of challenges that need to be addressed.

KEY WORDS:
artificial intelligence, computer-aided drug design, databases, machine learning

1 UVOD

Odkrivanje novih učinkovin je dolgotrajen postopek z visokimi finančnimi vložki. Začne se z izbiro in pripravo struktur tarč, iskanjem spojin zadetkov, njihovo optimizacijo do spojin vodnic, vrednotenjem v predkliničnih raziskavah in konča s potrditvijo varnosti in učinkovitosti v kliničnih raziskavah. V preteklih desetletjih za namen zmanjšanja stroškov in hitrejšega razvoja vse pogosteje uporabljamo računalniško podprto načrtovanje učinkovin (1). Zadnji napredek predstavlja vpeljava umetne inteligence (*artificial intelligence*, AI) v proces odkrivanja učinkovin (2). Priča smo razmahu umetne inteligence na številnih področjih, predvsem zaradi razvoja različnih algoritmov umetne inteligence, izboljšav strojne opreme (npr. grafične procesne enote, tensorska jedra, vzporedno programiranje) in dostopnosti vse večjih količin podatkov (3). O široki uporabnosti umetne inteligence med drugim pričajo

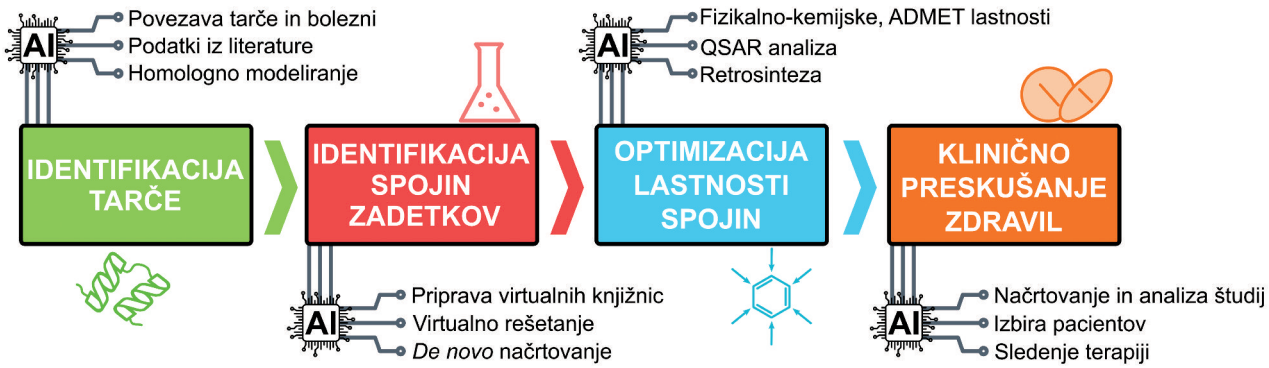
dosežki na področju prepoznave slik in zvoka, procesiranju naravnega jezika in igranju miselnih iger (2).

Tri ključne lastnosti, ki definirajo umetno inteligenco, so reševanje problemov, učenje iz izkušenj in prilagajanje novim izzivom (posploševanje). Za razliko od splošne umetne inteligence ozko umetno inteligenco uporabljamo samo za točno določene naloge, ki jim je namenjena. Strojno učenje je področje umetne inteligence, ki se ukvarja s programiranjem računalnikov, da se lahko učijo iz podatkov in v njih iščejo vzorce. Ena izmed novejših metod je globoko učenje, ki po navdihu človeških možganov za analizo podatkov uporablja večslojne nevronske mreže (4).

V naslednjih poglavjih bomo predstavili različne faze racionalnega odkrivanja novih učinkovin, v katerih se v določenih korakih že vključuje uporaba umetne inteligence (preglednica 1, slika 1).

Preglednica 1: Primeri uporabe umetne inteligence pri odkrivanju novih zdravil.
Table 1: Examples of using artificial intelligence in drug discovery.

Identifikacija tarče	**Priprava virtualnih knjižnic spojin**	**Racionalno načrtovanje spojin zadetkov**	**Optimizacija lastnosti spojin**	**Klinično preskušanje zdravil**
Pridobivanje podatkov iz literature	Priprava manjših usmerjenih in raznolikih knjižnic spojin	Virtualno rešetanje na osnovi strukture tarče: cenilne funkcije, ponovno ocenjevanje vezavnih konformacij	Priprava knjižnice analogov z določenimi fizikalno-kemijskimi lastnostmi	Pridobivanje podatkov iz literature
Povezave med genskim zapisom, proteini in nastankom bolezni	Odkrivanje novih ali manj zastopanih virtualnih kemijskih prostorov	Virtualno rešetanje na osnovi strukture liganda: QSAR, razvrščanje	Priprava knjižnice analogov na podlagi QSAR analize	Izbira pacientov
Repozicioniranje zdravil	Sestava knjižnic s specifičnimi lastnosti spojin	*De novo* načrtovanje z večparametrsko optimizacijo lastnosti spojin	Priprava in vrednotenje sinteznih poti za nove analoge	Spremljanje terapije
Priprava 3D strukture tarče			Napovedovanje lastnosti ADMET	Načrtovanje in analiza kliničnih raziskav



Slika 1: Uporaba umetne inteligence pri različnih fazah odkrivanja zdravilnih učinkovin.
Figure 1: Use of artificial intelligence at different stages of drug discovery.

2 VHODNI PODATKI

Vhodni podatki so ključni za pripravo modelov, ki omogočajo predvidevanje, ne glede na to, kateri algoritem uporabimo (5). Glavni izziv predstavlja pridobivanje podatkov v zadostni količini, ki so hkrati ustrezne kakovosti. Na področju odkrivanja novih učinkovin so na voljo zelo majhne količine podatkov v primerjavi z ogromnimi količinami informacij, ki jih uporabljajo v tehnologiji prepoznave slik ali procesiranju naravnega besedila (4). Do večjih količin lahko pridemo z združevanjem rezultatov iz različnih laboratorijev. Na ta način sicer povečamo variabilnost, saj v kompleksnih bioloških sistemih pod drugačnimi eksperimentalnimi pogoji pogosto dobimo drugačne ali celo nasprotujoče si rezultate (6). Če je model zgrajen na nezanesljivih podatkih, sta vprašljivi njegovi uporabnost in praktična vrednost (4). Vse večje količine rezultatov različnih bioloških testov lahko najdemo v prosto dostopnih podatkovnih zbirkah, kot sta ChEMBL (7) in PubChem (8). Veliko podatkov pa ostaja v lasti farmacevtskih družb, saj predstavljajo konkurenčno prednost (4). Možna rešitev je uporaba algoritmov umetne inteligence, ki jim zadostujejo že manjše količine podatkov (9, 10).

V procesu optimizacije učinkovine je potrebno identificirati spojino, ki deluje na izbrano tarčo, ima ustrezne farmakokinetične lastnosti in ne povzroča neželenih učinkov. Za takšno večparametrsko optimizacijo so potrebni številni testi, ki si sledijo od preprostejših do vedno zahtevnejših. Zahtevnejše raziskave izvajamo samo na najobetavnejših spojinah, zaradi česar je v podatkovnih zbirkah prisotnih veliko manjkajočih vrednosti, ki niso naključno razporejene po kemijskem prostoru. Poleg tega so podatki pogosto neuravnoteženi in vsebujejo veliko neaktivnih in malo aktivnih spojin ali obratno. Podatki iz rešetanj visoke zmogljivosti vsebujejo visok delež neaktivnih spojin, nasprotno pa je pri uporabi podatkov iz znanstvene literature problem redko poročanje negativnih rezultatov (4, 11).

Končni cilj uporabe umetne inteligence je predvideti obnašanje spojin v človeškem organizmu, za kar so na voljo najbolj omejene količine podatkov. Za takšne napovedi lahko uporabljamo nadomestne izide, npr. podatke, pridobljene na živalskih modelih, uporabljamo za napovedovanje kliničnih izidov, ker je količina takšnih raziskav veliko večja od količine raziskav na ljudeh (12). Podobno toksičnost spojin napovedujemo iz *in vitro* podatkov, kar zmanjša potrebe po testiranju na živalih (13). Tovrstni modeli imajo določene omejitve, saj povezave med nadomestnimi in končnimi izidi niso popolnoma pojasnjene (4).

3 IDENTIFIKACIJA TARČE IN PRIPRAVA NJENE STRUKTURE

Iskanje nove zdravilne učinkovine se pogosto začne pri hipotezi, da modulacija določene tarče v organizmu lahko zavre ali omili bolezensko stanje. Pri tem najpogosteje z literaturnimi viri najprej identificiramo tarčo, ki je povezana z nastankom bolezni. Zaradi vedno večje količine znanstvene literature se je pri iskanju teh povezav začela uporabljati umetna inteligenca, ki z napredkom v procesiranju naravnega jezika prečesava množice tekstovnih podatkov (*text mining*) iz različnih podatkovnih zbirk (14). Poleg literature pa umetna inteligenca uporablja tudi podatke o povezavah med proteini in njihovim genskim zapisom. S tem pristopom iščemo različne kombinacije genske variabilnosti, ki so odgovorne za nastanek bolezni (15). Tako so npr. raziskovalci na podlagi podatkov o izražanju mRNA, mutacijah proteinov in protein-protein interakcijah s pomočjo metode umetne inteligence identificirali 122 tarč, vpletenih v nastanek raka. Od teh je bilo 69 tarč že znanih. Dve novi tarči so ovrednotili s peptidnimi zaviralci, ki so izkazali močno antiproliferativno delovanje na rakavih celicah (16). Poleg identifikacije tarče, ki je potencialno vpletena v potek bolezni, nadalje lahko uporabimo metode umetne inteligence za napovedovanje, ali je izbrana tarča sploh primerna za razvoj modulatorjev (*target drugability*) (14).

Dodatno se je umetna inteligenca začela uporabljati tudi za iskanje novih tarč za že znane zdravilne učinkovine. V tem primeru gre za repozicioniranje ali prenamembo zdravil (*drug repurposing*). V farmacevtski industriji tak pristop uporabljajo pri iskanju zdravil za redke bolezni, saj se tako izognejo visokim stroškom razvoja nove zdravilne učinkovine. Iskanje novih tarč pa je industriji v interesu tudi za podaljšanje patenta zdravilnih učinkovin. Z umetno inteligenco najprej preiščejo literaturne podatke. Na podlagi množice kompleksnih vhodnih podatkov, ki predstavljajo velik informacijski šum, nato z algoritmi statistično ovrednotijo povezanost določenih zdravilnih učinkovin z različnimi tarčami. Pri tem uporabljajo različne povezave,

kot npr. stranski učinki zdravil, interakcije med zdravili, kemijske strukture zdravilnih učinkovin, interakcije proteinov in genski zapis (17, 18).

Včasih za tarčo za razvoj novih učinkovin ni znana njena trodimenzionalna struktura, ki bi omogočila strukturno podprti pristop načrtovanja spojin. V tem primeru pogosto uporabljamo homologno modeliranje. Gre za metodo, pri kateri strukturo tarčnega proteina predvidimo glede na znano aminokislinsko zaporedje, ki ga nato primerjamo z zaporedji proteinov z že znano strukturo. Implementacija umetne inteligence v razreševanje strukture proteinov je v zadnjih letih bistveno izboljšala natančnost napovedanih struktur (19). Orodje umetne inteligence, kot je npr. AlphaFold, se najprej uči na proteinih z znanimi strukturami. Na podlagi podatkov, kot so koti peptidnih vezi v proteinski strukturi in razdalje med aminokislinskimi pari, nato zgradi trodimenzionalni model proteina iz njegovega aminokislinskega zaporedja (20). Natančnejši modeli tako izboljšujejo nadaljnje korake v razvoju nove učinkovine.

4 PRIPRAVA VIRTUALNIH KNJIŽNIC SPOJIN

Glede na velikost ocenjenega virtualnega kemijskega prostora (~10^{63} molekul, ki upoštevajo pravila Lipinskega) je pred pričetkom virtualnega rešetanja (*virtual screening*, VS) smiselna priprava virtualnih knjižnic (21). Lahko pripravimo bolj usmerjene knjižnice s spojinami, za katere je že znana večja verjetnost za aktivnost na določenih tarčah, ali pa pripravimo manjše knjižnice z zelo raznolikimi spojinami. Velike podatkovne zbirke so lahko pri klasičnih pristopih časovno potratne in računsko zahtevne, z uporabo metod umetne inteligence pa lahko te težave zaobidemo (22).

S pomočjo nevronskih mrež je možno zmanjšati velike zbirke podatkov na manj kot odstotek njihove začetne velikosti. Kljub drastičnemu zmanjšanju je vseeno možno zajeti visoko reprezentativen vzorec spojin. Nasprotno lahko metode umetne inteligence razširijo kemijski prostor v del prostora, ki je manj zastopan v knjižnici, tako da pripravijo nove spojine (23). Primer tega je uporaba orodja AutoZoom, ki je raziskovalcem omogočilo razširitev svoje 1,7-milijonske virtualne knjižnice s približno 400.000 novimi spojinami. Te spojine so pridobili iz komercialno dostopne 8-milijonske knjižnice, kar bistveno poveča uspešnost njihove sinteze (24). Dodatno lahko prilagodimo knjižnico, da vsebuje le spojine z želenimi lastnostmi, npr. sintezno dostopnejše spojine (22). Takšna priprava virtualnih knjižnic omogoči, da virtualno rešetanje izvedemo na razumni časovni skali in hkrati povečamo možnost za odkritje biološko aktivne spojine. S pomočjo umetne inteligence lahko usmerimo iskanje po virtualnem kemijskem prostoru v del prostora, ki ne vsebuje patentiranih spojin.

5 RACIONALNO NAČRTOVANJE SPOJIN ZADETKOV

Za iskanje spojin zadetkov uporabljamo virtualno rešetanje kot racionalno alternativo rešetanju visoke zmogljivosti. Trenutno je glavna pomanjkljivost virtualnega rešetanja visok delež lažno pozitivnih rezultatov (1).

Prvi pristop je virtualno rešetanje na osnovi strukture tarče, pri katerem v postopku molekulskega sidranja iz virtualnih kemijskih knjižnic poiščemo komplementarne ligande. Program najprej generira vezavne konformacije (t. i. poze) ligandov v vezavnem mestu in jih nato s cenilno funkcijo razvrsti glede na predvideno aktivnost (1). V zadnjih nekaj letih za cenilno funkcijo razvijajo številne algoritme umetne inteligence, ki temeljijo na velikih količinah eksperimentalnih podatkov o interakcijah med ligandi in makromolekulami. V primerjalnih analizah dajejo boljše rezultate od klasičnih cenilnih funkcij, če jih primerjamo med seboj, pa nobena ne presega drugih v vseh vidikih. Izbira najprimernejšega algoritma umetne inteligence je zato odvisna od problema, ki ga obravnavamo (25). Umetno inteligenco uporabljamo tudi za obdelavo rezultatov sidranja s klasičnim programom, npr. s ponovnim ocenjevanjem vezavnih konformacij ali za izbor virtualnih zadetkov, ki tvorijo interakcije s ključnimi aminokislinskimi ostanki (26). Z uporabo modela za prepoznavanje aktivnih konformacij vScreenML so pripravili izbor in sintetizirali 23 potencialnih zaviralcev acetilholin-esteraze. Kar 21 spojin je zaviralo encim, od tega deset v nizkem mikromolarnem območju (26). Z naraščajočo količino eksperimentalnih podatkov lahko v prihodnosti pričakujemo uresničenje potenciala umetne inteligence na tem področju, kar bo nadomestilo ali izboljšalo klasični pristop rešetanja visoke zmogljivosti (2).

Drugi pristop, ki zajema večji delež uporabe umetne inteligence, je virtualno rešetanje na osnovi liganda. Kvantitativno raziskovanje odnosa med strukturo in delovanjem oz. določeno lastnostjo (QSAR/QSPR,

quantitative structure-activity/property relationship) je eden izmed prvih primerov uporabe nadzorovanega strojnega učenja v razvoju novih učinkovin. Metodo so še dodatno izboljšali z uporabo večslojnih nevronskih mrež (27). S pomočjo razvrščanja (*clustering*) in metod, ki temeljijo na podobnosti, lahko znanim aktivnim spojinam poiščemo strukturno podobne spojine za nadaljnje eksperimentalne raziskave. Primerjava različnih modelov umetne inteligence med seboj je težavna zaradi pomanjkanja primerjalnih analiz, ki uporabljajo iste nize podatkov (2).

Alternativa virtualnemu rešetanju je *de novo* načrtovanje, pri katerem na podlagi strukturnih podatkov o vezavnem mestu in znanja o tvorbi kemijskih vezi zgradimo nove, še neznane spojine, brez uporabe virtualnih knjižnic. Poleg vezave na tarčo je zaželeno, da imajo spojine ustrezen varnostni profil, ustrezne farmakokinetične lastnosti in jih lahko patentiramo. Tako lahko bolj učinkovito preiščemo samo določen del sicer ogromnega kemijskega prostora. Z uporabo umetne inteligence so na ta način že odkrili nove sintezno dostopne spojine. Razvoj gre še korak dlje, v načrtovanje učinkovin naslednje generacije, ki bodo hkrati delovale na več izbranih tarč. Gre za zahtevno večparametrsko optimizacijo, za katero je najbolj primerna uporaba umetne inteligence (28, 29).

6 OPTIMIZACIJA LASTNOSTI SPOJIN

Po odkritju spojine zadetka se začne kemijska modifikacija le-tega z namenom izboljšanja aktivnosti, selektivnosti ter različnih farmakokinetičnih in toksikoloških lastnosti (30). S pomočjo metod umetne inteligence lahko pripravimo knjižnice različnih analogov spojine zadetka, glede na njihove fizikalno-kemijske lastnosti in analizo QSAR (27). Umetna inteligenca je v pomoč tudi pri načrtovanju sinteznih poti, pri čemer pripravi retrosintezno analizo analogov do komercialno dostopnih začetnih sinteznih gradnikov (31). Pri tem preračuna različne možne sintezne poti glede na predvidene izkoristke, kompleksnost in ceno posamezne poti. Na podlagi teh informacij določi prioriteto za sintezo spojin. Kljub veliki količini podatkov o različnih kemijskih reakcijah, ima umetna inteligenca v tem primeru še veliko težav zaradi same kompleksnosti organske kemije, predvsem v primeru stereokemije, stranskih produktov ter množice različnih reagentov, katalizatorjev in reakcijskih pogojev (2, 32).

Vrednotenje lastnosti ADMET (*absorption, distribution, metabolism, elimination, toxicity*) z uporabo umetne inteligence lahko prav tako pomaga pri izbiri obetavnih spojin in zmanjša časovno-stroškovne vložke, povezane z raziskavami *in vitro* in *in vivo*. Tako umetno inteligenco uporabljamo za napoved lastnosti, kot so logP, pKa, temperatura tališča, topnost in intrinzična permeabilnost. Pristop globokega učenja izkazuje bistveno izboljšanje napovedne moči za topnost, pri intrinzični permeabilnosti pa predstavljajo oviro izlivne črpalke (*efflux pumps*). Na podlagi modelov številnih molekulskih deskriptorjev napovemo absorpcijo, distribucijo, metabolizem in eliminacijo (2). Toksikološke lastnosti spojin lahko napovemo s strukturno podprtim pristopom na tarče, kot so CYP450 (citokrom P450), hERG (*human ether-à-go-go-related Gene*), PXR (pregnanski X-receptor) in UGT (UDP-glukuronaziltransferaza). Pristop umetne inteligence pri pripravi toksikoloških modelov izkorišča spojine z znano aktivnostjo na teh tarčah (33). Za napovedovanje toksičnosti spojin uporabljamo tudi pristop na podlagi ligandov. Dober primer se je pokazal na tekmovanju *Tox21 Data Challenge*, kjer so tekmovalci pripravili modele za napovedovanje toksičnosti spojin. Na podlagi niza 10.000 spojin z znanimi toksikološkimi lastnostmi so s pomočjo metod umetne inteligence napovedali toksičnost 647 spojin. Izkazalo se je, da ima metoda globokega učenje bistveno boljšo napovedno moč kot druge metode (34).

7 KLINIČNO PRESKUŠANJE ZDRAVIL

Največji stroški v razvoju novih učinkovin nastajajo zaradi neuspehov v kliničnih raziskavah, saj samo eden od desetih kliničnih kandidatov uspešno pride na tržišče (35). Ker je na voljo vedno več digitalnih zdravstvenih podatkov, lahko z uporabo umetne inteligence skrajšamo dolgotrajne klinične raziskave in povečamo možnosti za uspeh. S tehnologijo procesiranja naravnega jezika lahko dodatne podatke pridobimo iz nestrukturiranih znanstvenih člankov in preteklih kliničnih raziskav. Zaradi stroge regulacije farmacevtske industrije je proces vpeljave novih tehnologij postopen, najprej pa je potrebno dokazati prednosti v primerjavi z obstoječo tehnologijo. Glavni pomisleki v zvezi z umetno inteligenco se pojavljajo zaradi slabega vpogleda v mehanizem odločanja njenih algoritmov (36).

S pomočjo umetne inteligence lahko analiziramo genomske in okoljske determinante pacientov in glede na izražanje tarče izberemo najbolj ustrezne paciente za drugo in tretjo fazo kliničnih preskušanj. Iskanje pacientov trenutno zahteva približno eno tretjino časa trajanja kliničnih raziskav. Po drugi strani lahko glede na izbrano populacijo določimo, kateri klinični kandidati imajo večje možnosti za uspeh. Med potekom kliničnih raziskav lahko z mobilnimi aplikacijami in nosljivimi napravami še natančneje spremljamo paciente in preprečimo izpad (*dropout*) ter povečamo adherenco (36, 37). S pomočjo mobilne aplikacije AiCure so spremljali jemanje zdravil pacientov s shizofrenijo in povečali adherenco za 25 % (38). Velik potencial se skriva v podatkih iz preteklih kliničnih raziskav, zlasti neuspešnih. Z umetno inteligenco bi lahko analizirali povezave med načrtovanjem kliničnih raziskav in njihovo uspešnostjo ter nove ugotovitve uporabili za boljše načrtovanje prihodnjih raziskav (36).

8 SKLEP

Uporaba umetne inteligence je smiselna takrat, ko z njeno pomočjo povečamo natančnost modelov *in silico* ter tako pospešimo razvoj zdravilnih učinkovin ali zmanjšamo njegove stroške. Pri tem ne obstaja le en vsesplošno uporaben pristop ali najboljši algoritem, temveč je potrebna uporaba različnih metod s kombinacijami podatkov, nato pa na podlagi validacije modelov izberemo najboljšega za dani problem. Zaradi uporabe različnih nizov podatkov je težavna medsebojna primerjava modelov, tudi če gre za reševanje istega problema. Za uporabnike in razvijalce bi bilo koristno, da standardiziramo validacijo modelov na neodvisnem nizu podatkov (2).

Kljub ogromnemu številu primerov uporabe umetne inteligence v literaturi se pojavljajo deljena mnenja glede njene uporabe za odkrivanje novih učinkovin: nekateri so prepričani o številnih priložnostih, drugi so skeptični in čakajo trdne dokaze (4). Potencial se kaže na številnih področjih odkrivanja novih učinkovin. Kot velja za vse nove koncepte, tudi uporaba umetne inteligence ni panaceja, ampak naj služi kot dodatno orodje za pomoč raziskovalcem, da bodo prišli hitreje do novih in boljših zdravil. Še vedno ostaja umetna inteligenca kot črna skrinjica, v katero nimajo natančnega vpogleda niti njeni snovalci.

9 LITERATURA

1. *Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov. 2010 Apr;9(4):273–6.*
2. *Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. Chem Rev. 2019 Sep 25;119(18):10520–94.*
3. *Gawehn E, Hiss JA, Brown JB, Schneider G. Advancing drug discovery via GPU-based deep learning. Expert Opin Drug Discov. 2018 Jul 3;13(7):579–82.*
4. *Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020 May;19(5):353–64.*
5. *Lean Yu, Shouyang Wang, Lai KK. An integrated data preparation scheme for neural network data analysis. IEEE Trans Knowl Data Eng. 2006 Feb;18(2):217–30.*
6. *Kalliokoski T, Kramer C, Vulpetti A, Gedeck P. Comparability of Mixed IC50 Data – A Statistical Analysis. PLOS ONE. 2013 Apr 16;8(4):e61007.*
7. *Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017 Jan 4;45(D1):D945–54.*
8. *Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019 Jan 8;47(D1):D1102–9.*
9. *Altae-Tran H, Ramsundar B, Pappu AS, Pande V. Low Data Drug Discovery with One-Shot Learning. ACS Cent Sci. 2017 Apr 26;3(4):283–93.*
10. *Baskin II. Is one-shot learning a viable option in drug discovery? Expert Opin Drug Discov. 2019 Jul 3;14(7):601–3.*
11. *Zakharov AV, Peach ML, Sitzmann M, Nicklaus MC. QSAR Modeling of Imbalanced High-Throughput Screening Data in PubChem. J Chem Inf Model. 2014 Mar 24;54(3):705–12.*
12. *Gorelick FS, Lerch MM. Do Animal Models of Acute Pancreatitis Reproduce Human Disease? Cell Mol Gastroenterol Hepatol. 2017 Sep 1;4(2):251–62.*
13. *Knudsen TB, Keller DA, Sander M, Carney EW, Doerrer NG, Eaton DL, et al. FutureTox II: In vitro Data and In Silico Models for Predictive Toxicology. Toxicol Sci. 2015 Feb 1;143(2):256–67.*
14. *Hameduh T, Haddad Y, Adam V, Heger Z. Homology modeling in the time of collective and artificial intelligence. Comput Struct Biotechnol J. 2020 Jan 1;18:3494–506.*
15. *Oprea TI, Bologa CG, Brunak S, Campbell A, Gan GN, Gaulton A, et al. Unexplored therapeutic opportunities in the human genome. Nat Rev Drug Discov. 2018 May;17(5):317–32.*
16. *Jeon J, Nim S, Teyra J, Datti A, Wrana JL, Sidhu SS, et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. Genome Med [Internet]. 2014 Jul 30 [cited 2021 Feb 7];6(7). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4143549/*
17. *Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019 Jun;18(6):463–77.*
18. *Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational Drug Repositioning: From Data to Therapeutics. Clin Pharmacol Ther. 2013;93(4):335–41.*
19. *Zhavoronkov A, Vanhaelen Q, Oprea TI. Will Artificial Intelligence for Drug Discovery Impact Clinical Pharmacology? Clin Pharmacol Ther. 2020 Apr;107(4):780–5.*

20. *Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020 Jan;577(7792):706–10.*
21. *Reymond J-L. The Chemical Space Project. Acc Chem Res. 2015 Mar 17;48(3):722–30.*
22. *Brown N, Ertl P, Lewis R, Luksch T, Reker D, Schneider N. Artificial intelligence in chemistry and drug design. J Comput Aided Mol Des. 2020 Jul;34(7):709–15.*
23. *Arús-Pous J, Blaschke T, Ulander S, Reymond J-L, Chen H, Engkvist O. Exploring the GDB-13 chemical space using deep generative models. J Cheminformatics. 2019 Mar 12;11(1):20.*
24. *Lin A, Beck B, Horvath D, Marcou G, Varnek A. Diversifying chemical libraries with generative topographic mapping. J Comput Aided Mol Des. 2020 Jul 1;34(7):805–15.*
25. *Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. WIREs Comput Mol Sci. 2015;5(6):405–24.*
26. *Adeshina YO, Deeds EJ, Karanicolas J. Machine learning classification can reduce false positives in structure-based virtual screening. Proc Natl Acad Sci. 2020 Aug 4;117(31):18477–88.*
27. *Winkler DA, Le TC. Performance of Deep and Shallow Neural Networks, the Universal Approximation Theorem, Activity Cliffs, and QSAR. Mol Inform. 2017;36(1–2):1600118.*
28. *Schneider P, Schneider G. De Novo Design at the Edge of Chaos. J Med Chem. 2016 May 12;59(9):4077–86.*
29. *Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. Nat Biotechnol. 2009 Feb;27(2):157–67.*
30. *Jiménez-Luna J, Pérez-Benito L, Martínez-Rosell G, Sciabola S, Torella R, Tresadern G, et al. DeltaDelta neural networks for lead optimization of small molecule potency. Chem Sci. 2019 Dec 4;10(47):10911–8.*
31. *Segler MHS, Waller MP. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. Chem – Eur J. 2017;23(25):5966–71.*
32. *Szymkuc S, Gajewska EP, Klucznik T, Molga K, Dittwald P, Startek M, et al. Computer-Assisted Synthetic Planning: The End of the Beginning. Angew Chem Int Ed. 2016;34.*
33. *Göller AH, Kuhnke L, Montanari F, Bonin A, Schneckener S, ter Laak A, et al. Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. Drug Discov Today. 2020 Sep 1;25(9):1702–9.*
34. *Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: Toxicity Prediction using Deep Learning. Front Environ Sci [Internet]. 2016 [cited 2021 Jan 21];3. Available from: https://www.frontiersin.org/articles/10.3389/fenvs.2015.00080/full*
35. *Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014 Jan;32(1):40–51.*
36. *Harrer S, Shah P, Antony B, Hu J. Artificial Intelligence for Clinical Trial Design. Trends Pharmacol Sci. 2019 Aug 1;40(8):577–91.*
37. *Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. Contemp Clin Trials Commun. 2018 Sep 1;11:156–64.*
38. *Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, et al. Use of a Novel Artificial Intelligence Platform on Mobile Devices to Assess Dosing Compliance in a Phase 2 Clinical Trial in Subjects With Schizophrenia. JMIR MHealth UHealth. 2017 Feb 21;5(2):e18.*